DOCUMENT RESUME

ED 247 395 CE 039 407

AUTHOR Thornton, Carl L.
TITLE The Four Dimensions of Course Evaluation Data: Analysis Leads to More Effective Management Decisions.
PUB DATE May 84
NOTE 22p.; Summary of a presentation at the National Convention of the American Society for Training and Development (Dallas, TX, May 1984).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adult Education; *Course Evaluation; Data Processing; *Decision Making; *Evaluation Methods; *Evaluation Utilization; Job Training; *Labor Force Development; Staff Development; Summative Evaluation

ABSTRACT

Course evaluation data can be a meaningful resource to improve the human resource development (HRD) practitioner's decision making. Recommended properties of an effective evaluation form are separation of the overall evaluation from the evaluation of specific aspects; a one-to-nine rating scale; strong, explicit verbal anchors on the rating scale; sufficient items to reflect all aspects; and a printed form. Several case studies have been conducted to compare courses and highlight the kinds of analysis that can be made once a good form and an adequate database have been established. A factor analysis of data from over 600 respondents across 24 courses reveals four dimensions being measured by the course evaluation questionnaire. As a result of the factor analysis, the number of specific items has been shortened from 18 to 10: presentation, relevance to job, coverage of subject, quality of materials, value for meeting objectives, instructor's expertise/enthusiasm, level of difficulty, classroom facilities, participant quality, and implementability of ideas. Recommendations to HRD professionals to change their approach to course evaluation data are (1) evaluation form redesign, (2) database development, and (3) use of a computer program to do the number crunching, calculate the average and median, sort data, and convert numbers to graphics. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247395

The Four Dimensions of Course Evaluation Data: Analysis Leads to More Effective Management Decisions

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Carl L. Thornton

NCR Corporation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D Lund

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Don Kirkpatrick (1967) pointed out that evaluations at the end of a course are not enough. Organizations must determine if there are improved results back on the job. Today, this is an "obvious truth" that is well known to HRD practitioners. However, this knowledge should not preclude the accurate gathering of course evaluation data. Such data need not be an object of derision, but, rather, it can be a meaningful resource to improve your decision making regarding courses.

My own research over the past several years has led me to believe that course evaluations are much more than a "smiles test." Adult learners have more than a warm glow (or heartburn) following an educational experience. They do not generate one giant halo effect when they evaluate a program.

To demonstrate how I reached this conclusion, I would like to present some concepts that I believe are important in constructing an evaluation form. Then, I would like to describe some case studies of course comparisons with supporting statistics.

CE039407

Psychometric Properties of an Effective Evaluation Form

1. Separate the overall evaluation from the evaluation of other aspects of the course. This reduces the halo effect.

I accomplished this by placing the overall evaluation on the front page, followed by open ended questions regarding training needs and suggestions for future programs. Eighteen specific ratings of the course were placed on the back of the form. Hopefully, this broke up the respondent's mental set and got their global feeling about the course out the way before making specific evaluations.

2. Use a one-to-nine rating scale.

   This allows greater differentiation where such can be made. The commonly used one-to-five scale lacks reliability of measurement; distinctions cannot be made, where they might actually exist. At the other extreme, scale values above nine add little or no increased accuracy in the numbers produced. Furthermore, they can lead to gross clerical errors when data is analyzed. For example, reversing digits and recording a 90 for an 09 is far more damaging than mistaking a 7 for a 9.

3. Use strong, explicit verbal anchors on the rating scale.

   a. Average can be interpreted as mediocre. "Typical, expected level" is a better choice of terms. This implies competence instead of mediocrity and provides a more stable midpoint.

   b. The high end should reflect the very highest level of attainment, such as "the best" or "one of the very best" that has been experienced. This reduces the tendency to be overly lenient by restricting the use of the high end of the scale to truely outstanding programs.

   c. Conversely, the low end should reflect extremely poor performance.

4. Use enough items to reflect all aspects of the course you wish to evaluate.

   Later, I will present 18 items along with the results using these, followed by a shorter list of 10 items. This reduced list was obtained after data analysis warranted such a reduction.

5. Take your evaluation form to the printer.

   Although a small matter, the professional appearance of your evaluation form does convey the fact that careful evaluation is important. Although I can't prove it with statistics, I think the appearance of the form conveys the professionalism you wish to establish, and the importance of the data you are asking for.

Course Comparisons

Next, I would like to present several case studies which highlight the kinds of analysis that you can make, once a good form and an adequate data base has been established.

Case #1

1. Same Course Title - Product Management
2. Same Length - Two Day Program
3. Two Different Instructors
4. Independent Course Preparation by each Instructor

In one class, the instructor was a college professor who taught the fundamental concepts of Product Management. In the other class, the instructor was a consultant who previously had worked in the Product Management function for a company, and subsequently set up his own consulting firm to assist companies that wanted to set-up or improve the Product Management function.

Table one shows the results of the evaluations from the two classes. Note that differences in the two classes cannot be attributed to differences in expertise or instructor enthusiasm. Both received nearly identical ratings on these two measures. Would you hazard a guess as to which one received the higher rating? Of course, your answer may reveal more about your attitudes regarding professors ("those who can't do, teach"?) or consultants ("borrow your watch - tell you what time it is - then keep your watch"?)

Notice, that the differences occurred despite not only identical, but very high ratings of expertise and enthusiasm. Also, notice that the large differences reflected items that appear to be tapping the practical value or utility of the presented material. In this case, the consultant was perceived as producing a course of more practical value to participants. (Caution: don't generalize to other consultants or academicians; collect your own data to verify your comparisons.) As a result of this difference, the consultant was chosen for future course offerings a with a subsequently favorable influence on the organization. Some groups even invited the consultant to extended sessions to assist with particular product plans as a result of the perceived value of the two-day workshop.

Case #2

1. Same Course Title: Finance for the Non-Financial Manager
2. Same Length: Four Day Program
3. Same Instructor's Manual with Structured Participant Exercises and Case Studies
4. Two instructors: one was brand new to the course and the company; one had taught the course several times over the last

two years.

Table two shows the results. Here, it is very easy to pick out the differences. Simply look across from "instructor expertise" in the table. In my first example, Case #1, there was no difference; in Case #2, there is a huge difference. That's the bad news. The good news is that the gaps in other areas are <u>smaller</u> than they were in Case #1. My interpretation is that the upfront course development work with the instructor's manual paid off. The course was "robust" enough to allow the novice to achieve the intended results.

Parenthetically, these differences were <u>not</u> apparent by looking at columns of numbers. The tendency was to look at the one poor result (a low rating of expertise) and initially make the decision to reject the instructor outright. Putting the results into a profile helped provide a more balanced view. - This leads to an important suggestion:

<u>Always</u> convert numbers to some graphic form or picture to improve your perspective and your decision making.

Case #3

1. Same Course Title: Improving Product Quality
2. Same Length: Two hours per week for 16 weeks
3. Same Training Package
4. Different Discussion Leaders
5. Different Audiences: Manufacturing Managers versus Managers of Computer Science Personnel

I call this: "the right course for the right audience versus the right course for the wrong audience." Table three shows the dramatic differences between the responses from the two audiences. Notice, that an overall negative halo does not

exist. Identical ratings occur where expected. Both groups used identical training materials, so, organization, coverage, quality of written materials as well as the quality of audiovisual materials show very similar results.

Five very large differences occur for the computer science personnel. In the training materials all examples and case studies came from Manufacturing. The computer people could not relate these examples to their environment.

When this data was being analyzed, two groups of managers in the computer area had procrastinated and apparently resisted the program. After analyzing the data in graphic form, and noting the wide gaps for this audience, two recommendations were offered:

1. don't force the program on groups where it won't be accepted or,
2. provide supplemental examples and case studies and instruct the discussion leader to continually show how the general principles apply to a computer programming environment. Failure to recognize these differences would have produced meaningless education, if corrections had not been made.

Case #4 and #5

1. Same Subject
2. Same Materials
3. Same Instructor
4. Same Types of Employees
5. Different Locations

For the statisticians reading this article, I have provided some evidence regarding

the consistency of results when all variables except location are held constant. In both cases, the locations were chosen which had the largest number of participants so that the choice would not be based on any biases on my part.

Table four shows the results from a course in Technical Writing taught at a manufacturing facility in Columbia, South Carolina and another manufacturing facility within the same company but located in Waterloo, Canada. At each location about 30 people evaluated the course. Notice, the striking similarity in results. Rarely, do the results differ by more than a few tenths of a point. Thus, this course and instructor have a clear, identifiable profile that is quite consistent from place-to-place. In short, the results are statitically reliable.

An even more dramatic example is offered in Table five. Here a sample of 90 participants from each of two locations (Cambridge, Ohio and San Diego, California) evaluated a course. Notice that the differences are even smaller, when the target populations are the same (manufacturing) and the sample sizes are larger (N=90). This comparison proves to be even more remarkable when I let you know that the the subject being profiled is the <u>same</u> one shown to have wide difference between groups in Table three.

Thus, table four and five give confidence in the accuracy and consistency of the data that is produced with an evaluation form when care is taken to construct it carefully.

Finally, I conducted one other analysis that clearly demonstrates that something more than halo or "gut feel" is involved in evaluations made by the adult learner. Taking data from more than 600 respondents across 24 courses, I utilized a technique called Factor Analysis to determine the actual number of dimensions being

measured by the questionnaire. As implied by the title of this article four such dimensions were found. Table six shows these results. The numbers in the table can range from zero to 1.0. Values below .35 were left out so that the table would be easier to read. The higher the number, the higher the correlation between the item and the hypothetical Factor (1-4) across the top of the table. Naming each factor is a matter of personal interpretation. Factor two might be termed "practical value" or "the substance" of the course, where Factor one is "the style". Clearly, both substance and style are necessary for good, overall evaluations. However, if your program has style but no substance your participants will tell you.

Forms Revision

As a result of the factor analysis, I shortened the number of specific items from 18 to 10. These are listed below:

1. Presentation/Delivery
2. Relevance to your job
3. Coverage, organization of the subject
4. Quality of written, audiovisual materials
5. Value for meeting your objectives
6. Instructor's expertise/enthusiasm
7. Level of difficulty/homework
8. Classroom facilities/meals, coffee breaks, etc.
9. Participant quality
10. Implementability of ideas back at work

By the way, I believe that item number 6, expertise and enthusiasm, are clearly separate entities in our minds, but, statistically, they "go together". I think the reason may go something like this: How do we know someone is an expert? They enthusiastically tell us so.

Final Recommendations

If this discourse regarding the value of course evaluations has been compelling, I recommend that you as an HRD professional change some behaviors regarding your approach to course evaluation data.

1. Redesign your evaluation form to make it the best possible.
2. Develop a data base, so that meaningful comparisons can be made.
3. Convert your number comparisons to graphics to gain a balanced perspective. Tables of numbers may impress upper management and show how hard you work, but they don't help you help them make important decisions. Relegate the pages of numbers to a technical appendix. This will show how hard you work, but it won't allow the numbers to get in the way.
4. If you work for a small organization, consider joining a network of firms that can form a common data base.
5. Enlist the aid of a computer program to do the number crunching. There's no honor in tedious hours of work that can be done quickly on the computer. One such program, Statistical Package for the Social Science (SPSS) can be found at most universities. Also, statistics programs are available for the personal computer found in many offices today.
6. Use a statistical package that can calculate both the average and the median. The median is a more stable and accurate number if only a few people disagreed greatly from the general opinion of the class.
7. Also, a program that can sort data by location or department codes is very helpful (SPSS can do #6 and #7).
8. Find a computer program to convert your numbers to graphics.

This is another time-saver. Also, whether we like it or not, people believe printouts more than they do hand drawn examples. Use this phenomena to your advanage.

9. Finally, brush up on your statistics. Often, however, your data will speak for itself.

I hope that this dialogue * will help you to improve your HRD-analytical competencies, and allow you and your management to make important, accurate, and timely decisions regarding your educational programs.

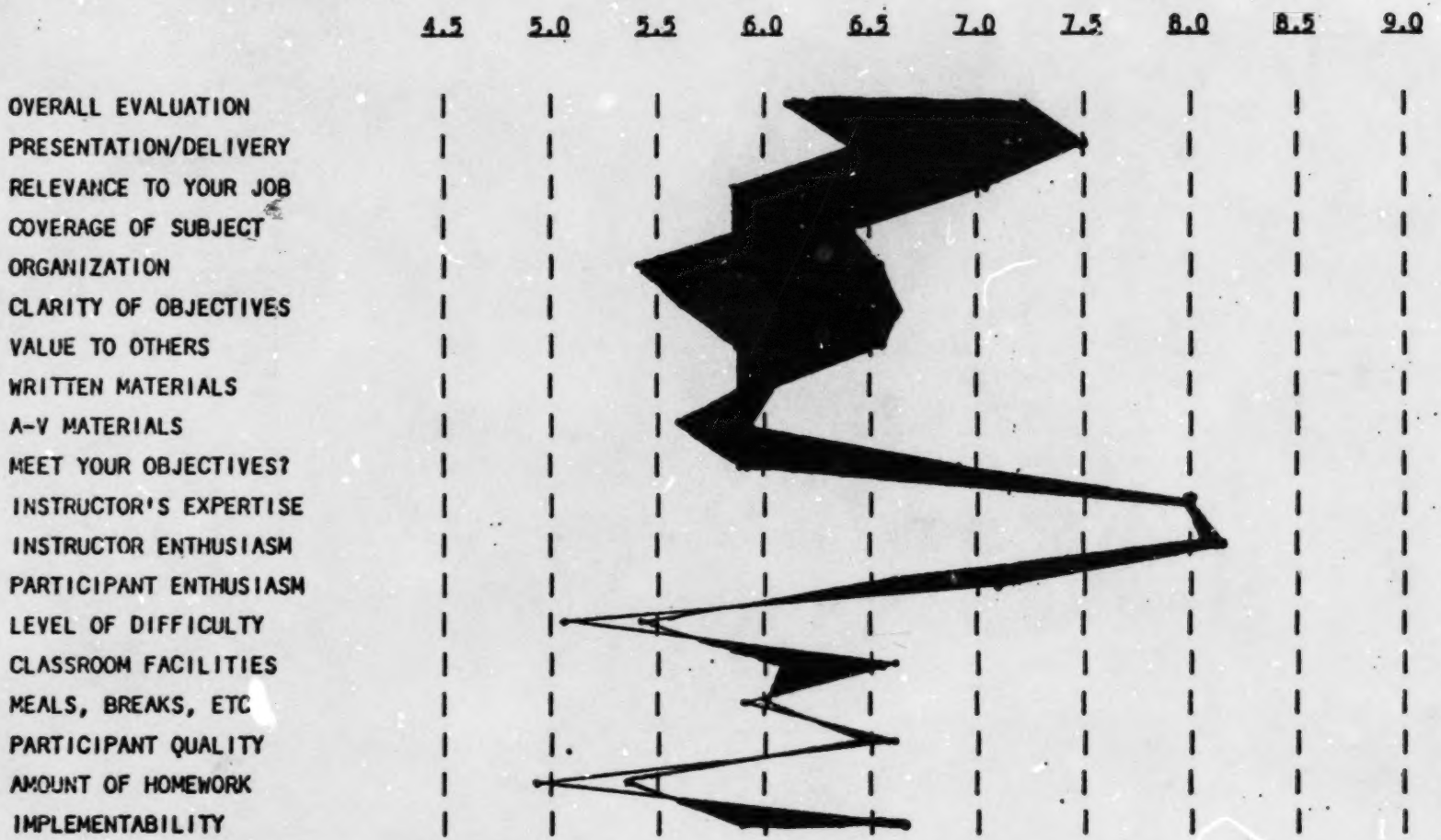
Reference

Kirkpatrick, D. L. "Evaluation of Training" in Craig, R. L., and Bittel (Eds.), Training and Development Handbook. New York: McGraw-Hill, 1967, 87-112.

* This article represents a summary of information presented at The ASTD National Convention - Dallas, 1984 with the title "Making the Most of Course Evaluation Data."

TABLE ONE. THE PROFESSOR AND THE CONSULTANT

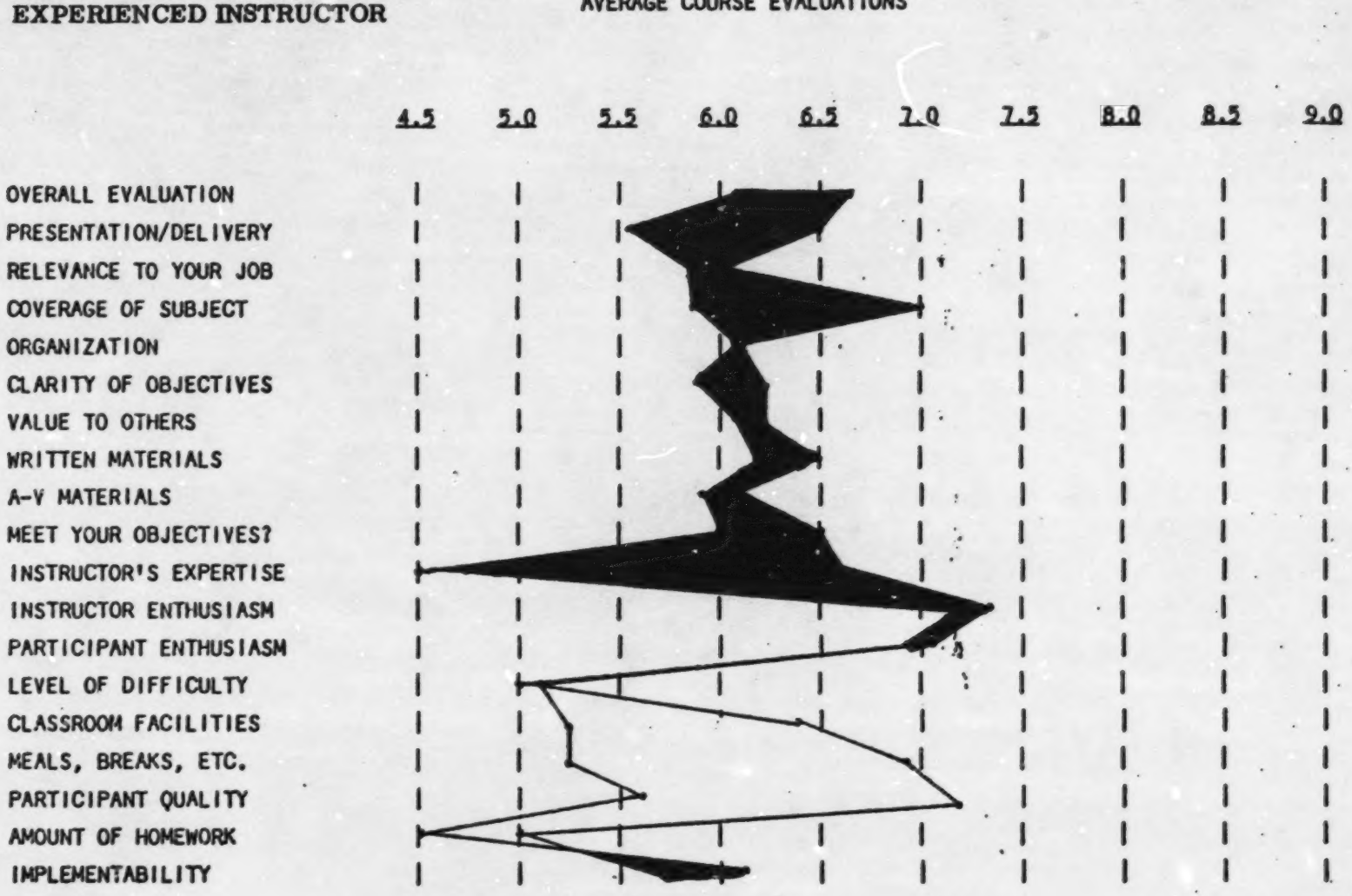
AVERAGE COURSE EVALUATIONS



* **The dark area represents the difference between the course rated highest overall and the lower rated course.**

White space between the lines indicates situations where the course rated lower overall had higher ratings.

TABLE TWO. THE NOVICE AND THE EXPERIENCED INSTRUCTOR



White space between the lines indicates situations where the course rated lower overall had higher ratings.

TABLE THREE. THE "RIGHT" AND "WRONG" AUDIENCE

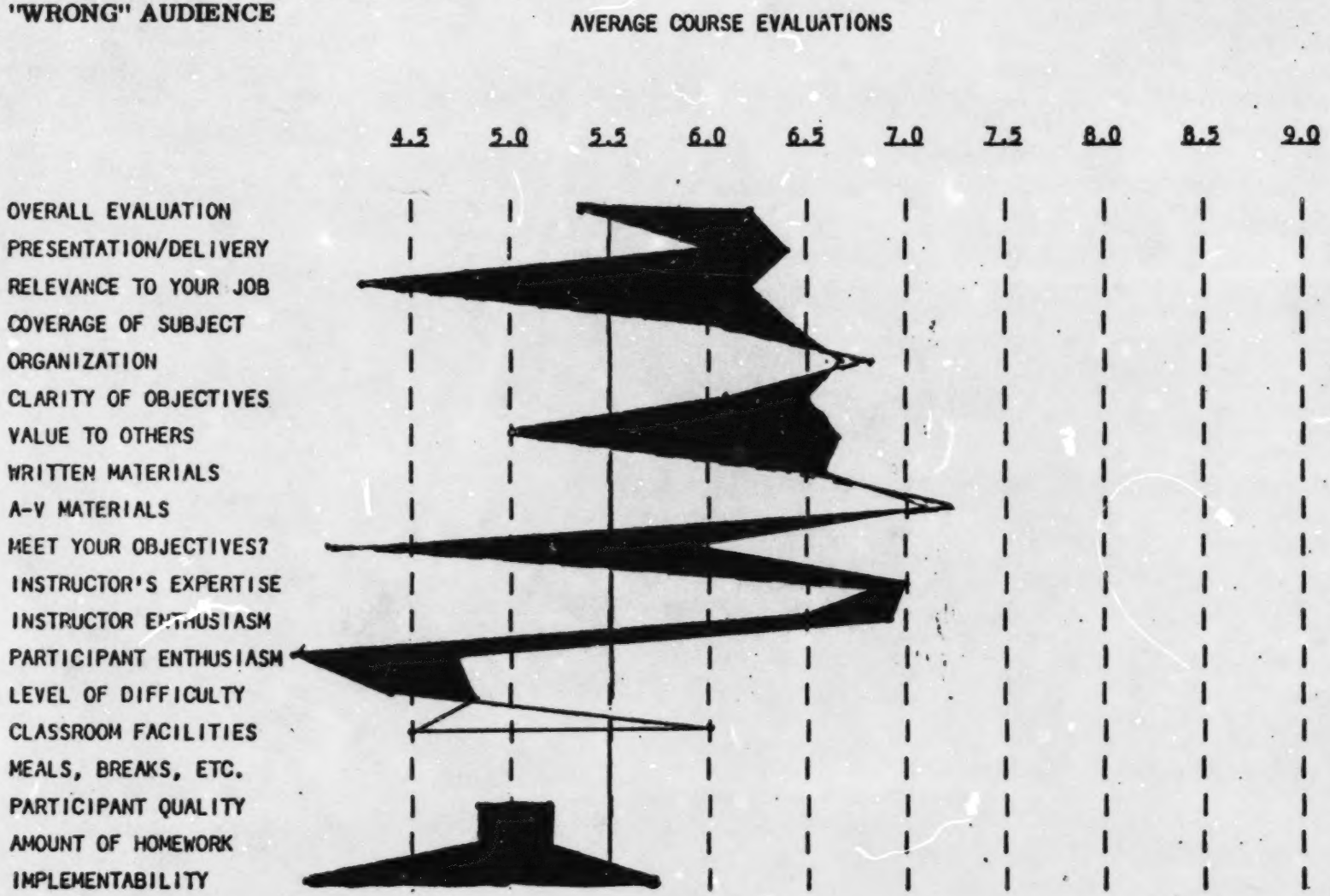


TABLE FOUR. DIFFERENT LOCATIONS: EVERYTHING ELSE THE SAME (N=30). AVERAGE COURSE EVALUATIONS

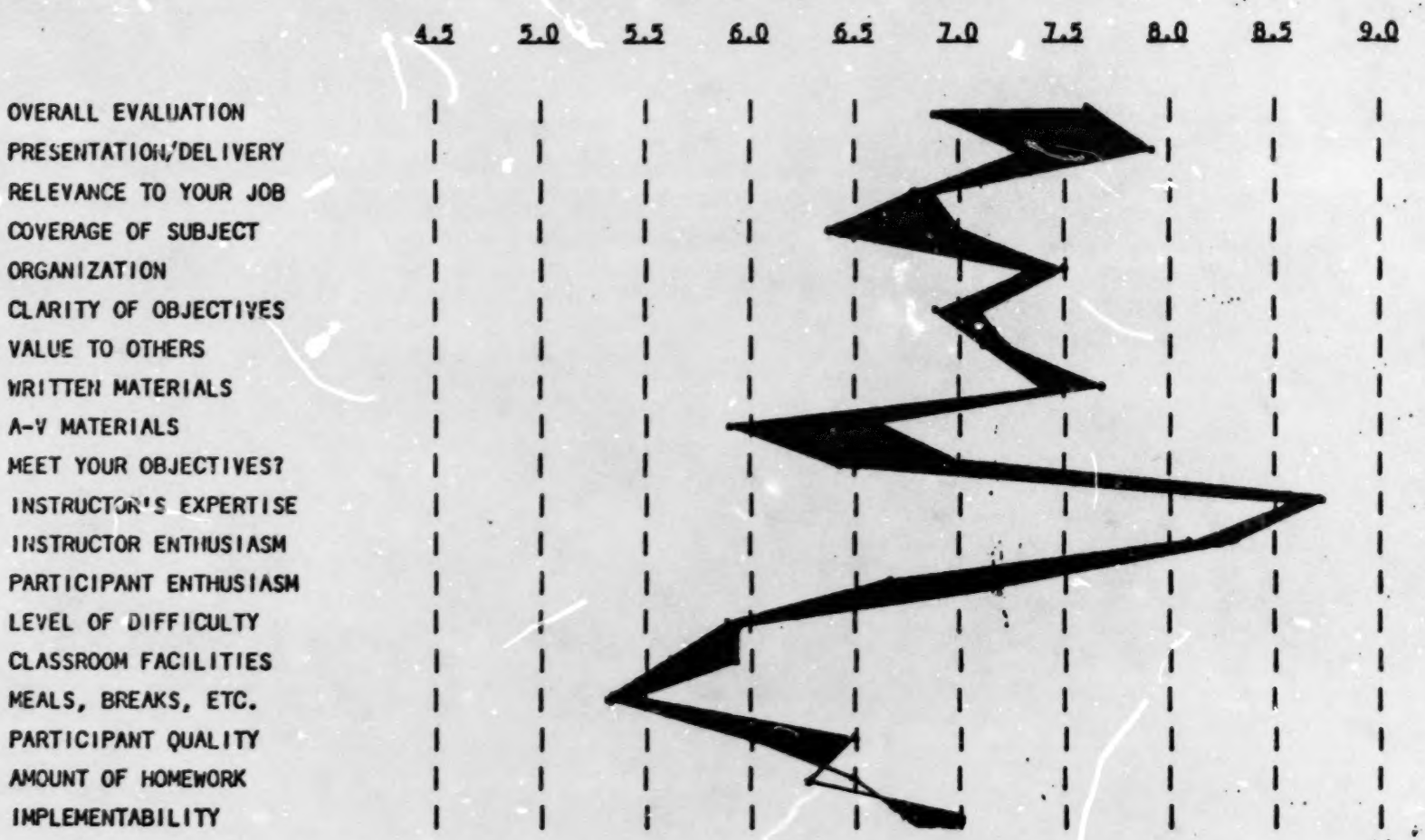


TABLE FIVE. DIFFERENT LOCATIONS: EVERYTHING ELSE THE SAME (N=90).

AVERAGE COURSE EVALUATIONS

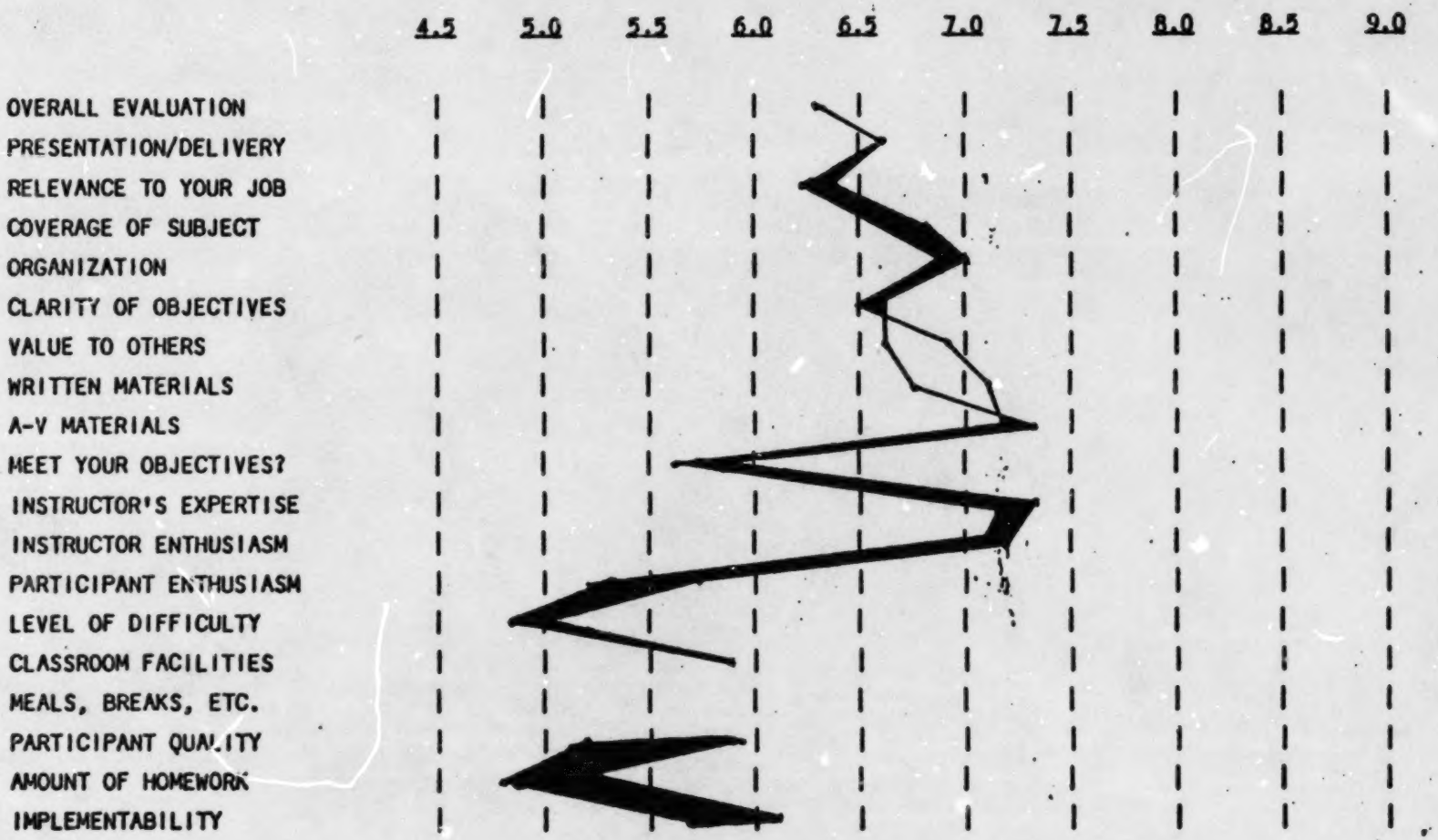


TABLE SIX. THE FOUR FACTORS IN THE COURSE EVALUATION DATA

	I	II	III	IV
OVERALL EVALUATION	.51	.47		
PRESENTATION/DELIVERY	.80			
RELEVANCE TO YOUR JOB		.78		
COVERAGE OF SUBJECT	.58	.46		
ORGANIZATION	.66	.38		
CLARITY OF OBJECTIVES	.47	.49		
VALUE TO OTHERS		.66		
WRITTEN MATERIALS	.39	.49		
A-V MATERIALS			.44	
MEET YOUR OBJECTIVES?		.78		
INSTRUCTOR'S EXPERTISE	.84			
INSTRUCTOR ENTHUSIASM	.83			
PARTICIPANT ENTHUSIASM	.57		.42	
LEVEL OF DIFFICULTY				.66
CLASSROOM FACILITIES			.67	
MEALS, BREAKS, ETC.			.76	
PARTICIPANT QUALITY			.70	
AMOUNT OF HOMEWORK				.73
IMPLEMENTABILITY		.80		

02/07/84:CLT:CLM:16